

Automate the business loan approval system for a central bank using historical data of borrowers

Problem Statement

About the data

The dataset is from the U.S. Small Business Administration (SBA)

The U.S. SBA was founded in 1953 on the principle of promoting and assisting small enterprises in the U.S. credit market (SBA Overview and History, US Small Business Administration (2015)). Small businesses have been a primary source of job creation in the United States; therefore, fostering small business formation and growth has social benefits by creating job opportunities and reducing unemployment.

There have been many success stories of start-ups receiving SBA loan guarantees such as FedEx and Apple Computer. However, there have also been stories of small businesses and/or start-ups that have defaulted on their SBA-guaranteed loans.

Business Objectives

This case study aims to identify patterns which indicate if a client has difficulty paying their instalments which may be used for taking actions such as denying the loan, reducing the amount of the loan, lending (to risky applicants) at a higher interest rate, etc. Automate a loan Approval system

Presented by
Jubin Benny

Overall Approach :

- ▶ **1. Understanding Domain Variables using Data Dictionary given:**

'sba_National.csv' is data dictionary which describes the meaning of the variables.

- ▶ **2.install PostgreSQL on your system and Initiate Pgadmin**

PostgreSQL is a powerful, open-source object-relational database system with over 35 years of active development that has earned it a strong reputation for reliability, feature robustness, and performance.

- ▶ **3. Importing the Data file**

Create your data table & import the file

- ▶ **4. Checking the structure of Data**

No of Columns & Rows present

Type of Data present i,e:int/float

Checking Numerical variables i,e:Count,Max,Min,std etc

► 5. Data Cleaning & Manipulation

Removing all the impurities such as \$ sign ,Letters in Numeric columns and converting all Continuous Columns to int

Deleting all the rows where value is null

► 6. Analysing Columns to Select the best Metrics for Loan Automation

MIS_Status : Loan status charged off = CHGOFF, Paid in full =PIF

LowDoc : LowDoc Loan Program: Y = Yes, N = No

Term :Loan term in months

► 7. Analysis

Writing Sql Queries to get insights From the data & using it to create an eqn to automate the new clients loan Approval or Denial

Traditional Loan Application Process

- ▶ Manual collection and submission of documents
- ▶ Lengthy paperwork and documentation requirements
- ▶ Time-consuming verification and processing
- ▶ Potential for errors and delays

Automated Loan Application Process

- ▶ Online Application and Document Upload
- ▶ Automated Document Verification
- ▶ Credit Scoring and Risk Assessment
- ▶ Automated Decision-Making: Based on the predefined rules and risk thresholds

Advantages of Automated Loan Approval System

- ▶ **Comprehensive Insights:** The queries cover a range of metrics, including loan approval rates, loan amounts, loan status distribution, industry analysis, and loan term analysis. These insights provide a holistic view of the loan portfolio, borrower behavior, and lending patterns.
- ▶ **Performance Evaluation:** By examining metrics such as loan approval rates, loan distribution by year, and loan status distribution, lenders can assess the performance and effectiveness of their lending practices. These insights can help identify areas of improvement and inform strategic decision-making.
- ▶ **Risk Assessment:** Queries related to charged-off principal amounts, loan status percentages, and loan term analysis provide valuable information for risk assessment. Lenders can identify potential risk factors, evaluate the creditworthiness of borrowers, and implement risk mitigation strategies.

- ▶ **Compliance Monitoring:** The queries can be extended to include compliance checks, ensuring adherence to regulatory requirements. By incorporating compliance metrics, lenders can mitigate risks associated with non-compliance and maintain regulatory standards.
- ▶ **Efficiency and Cost Savings:** Automated systems and advanced SQL queries enable faster data analysis, reducing the manual effort required for data extraction and manipulation. This improves operational efficiency, saves time, and reduces costs associated with manual processing.
- ▶ **Data-Driven Decision Making:** The queries provide data-driven insights that facilitate informed decision-making. By leveraging the metrics obtained from these queries, lenders can make more accurate assessments of loan applications, improve risk management, and optimize resource allocation.

Overall, the advantages of these queries lie in their ability to provide comprehensive insights, facilitate performance evaluation, support risk assessment, aid compliance monitoring, drive efficiency, and enable data-driven decision making in the loan approval process.

► Analysing Columns to Select the best Metrics for Loan Automation

	avg_loan_term numeric	avg_number_of_employees numeric	avg_jobs_created numeric	avg_jobs_retained numeric	num_revolving_credit bigint	num_low_doc_loans bigint	total_loans bigint	num_approved_loans bigint	num_disapproved_loans bigint
1	110.7731111442876975	11.4045628452438516	8.4056792126156198	10.7771265015809634	201385	110221	897554	157484	0

avg_loan_term: Average length of the loan in months

avg_number_of_employees: Average number of employees in the business

avg_jobs_created: Average number of jobs created by the loan

avg_jobs_retained: Average number of jobs retained by the loan

num_revolving_credit: Number of loans with a revolving line of credit

num_low_doc_loans: Number of loans through the LowDoc Loan Program

total_loans: Total number of loans in the dataset

num_approved_loans: Number of loans that were approved

num_disapproved_loans: Number of loans that were disapproved or charged off

The top 3 customers with highest bank balance who have not defaulted on a loan in the last 3 years.
This will help us understand your ideal loan applicants.

	name character varying	city character varying	balance_gross integer	mis_status character varying	chgoff_prngr integer	approvalfy1 integer	rank bigint
1	CURTIS JOHNSON GENL BLDG CONTR	BAKERSFIELD CALIFORNIA	0	CHGOFF	0	1969	1
2	RALPH O BURNETT	BAKERSTOWN	0	P I F	0	1969	2
3	KINGS INN	SEATTLE	0	P I F	0	2003	7

The percentage of defaults as per urban and rural status

	urbanrural integer	percentage_paid numeric	percentage_defaulted numeric
1	0	92.7878620140623	7.122577114087489
2	1	75.30213662864816	24.4061379204841
3	2	81.01711649278032	18.71422197328574

we can say that the deefaulters in urban are more compared to rural

The high scaled clients with loan amounts more than 50,00,000

	name character varying	gr_appv integer	mis_status character varying
1	Simon PLC Attorneys & Couns...	5472000	P I F
2	'The Sports Basement, Inc.'	5000000	P I F
3	NetRiver Int. LLC	5000000	P I F
4	International Feed.com Corpora	5000000	P I F
5	Telex Metals, LLC	5000000	P I F
6	McCarty Equipment, Inc	5000000	P I F
7	Premier Energy Services, Inc.	5000000	P I F
8	Yonkers Gateway Motel LLC d...	5000000	P I F
9	Stor-All Pontchartrain, LLC	5000000	P I F
10	Industrial Energy Applications	5000000	P I F
11	Ideal Aerosmith, Inc.	5000000	P I F
12	CARTER BROTHERS, LLC	5000000	P I F
13	Interco Trading Inc	5000000	P I F
14	Greenland (America) Inc.	5000000	P I F
15	Brisam Brooklyn LLC dba: Co...	5000000	P I F
16	Interco Trading, Inc.	5000000	P I F
17	Bentrani Watches, LLC	5000000	P I F
18	Metlcast Industries LLC	5000000	P I F
19	SEATTLE SAFETY, LLC	5000000	P I F
20	Central Woodwork, Inc.	5000000	P I F
21	Quest Aircraft Company, L.L.C.	5000000	P I F
22	CB Franchising LLC	5000000	P I F
Total rows: 41 of 41 Query complete 00:00:00.526			

we can say that if Gross amount of loan approved by bank is greater than 50,00,000 then there are 100% chances of loan getting paid in Full

Percentage of loan (paid and defaulted)application by credit score

	lowdoc character varying	percentage_charged_off numeric
1	N	18.686518338009133
2	Y	8.98558350949456

- ▶ 90% of the customers credit score is not good and the percentage of chgoff is 18% when compared to customers with good credit score is 8.9%

checking your credit standing using different metrics. You can differentiate between key figures and use different key figures according to your bank's needs.

	loan_status text	loan_count bigint
1	Approved	190915
2	Declined	706639

- ▶ Eg:
 - Gr_Appv is >50,000 & less than 2.5 lakhs
 - Term :<= 84 Months
 - ChgOff_PrinGr:<= 10,000
 - Disbursement_Gross: <= 2.5lakhs

Conclusions:

1. Princess Manufacturing co, Kings Inn, Curtis Johnson are the ideal loan applicants
2. we can say that if Gross amount of loan approved by bank is greater than 50,00,000 then there are 100% chances of loan getting paid in Full
3. The defaulter's in urban are more compared to rural.
4. 90% of the customers credit score is not good and the percentage of chgoff is 18% when compared to customers with good credit score is 8.9%
5. The Loan with term 40-60 have the most charged off Accounts
6. The gross approval amount was highest in 2011 from (1962-2014)
7. Top Defaulters highlights are
 1. "CHOICE VENDING INC" has taken loan of 24 Lakhs with 22lakhs 23 thousand chgoff amount, loan approved was in 1999
 2. "UNITED YARNS, CO. INC." has taken loan of 24 Lakhs with 21lakhs 57 thousand chgoff amount, loan approved was in 1999
 3. "DOLPHIN LANES OF SOUTH FLORIDA" has taken loan of 22 Lakhs with 17lakhs 98 thousand chgoff amount, loan approved was in 2000
8. These are the top three Industries with highest loan Amounts
 1. Accommodation and food services
 2. Health care and social assistance
 3. Retail trade

Thank You